NTNU – Trondheim
Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for **ST1201/ST6201 Statistical methods**

Academic contact during examination: Nikolai Ushakov

Phone: 45128897

Examination date: 04 December 2015

Examination time (from–to): 09:00 – 13:00

Permitted examination support material: C:
– Tabeller og formler i statistikk, Tapir forlag,
– K.Rottman. Matematisk formelsamling,
– Stamped yellow A4 sheet with your own handwritten notes,
– Calculator: HP30S, Citizen SR-270X, Citizen SR-270X College or Casio fx-82ES PLUS.

Other information:

Language: English

Number of pages: 4

Number pages enclosed: 0

Checked by:

Date Signature

Problem 1

Let X be χ^2-distributed with $2n$ degrees of freedom ($n > 2$).

a) Show that

$$\mathrm{E}(X^{-1}) = \frac{1}{2(n-1)} \quad \text{and} \quad \mathrm{E}(X^{-2}) = \frac{1}{4(n-1)(n-2)}.$$

Problem 2

The literature states that the mean length in Norway of the tail of a mammalian species is 30 cm. A biologist think that the mean (that is, the expected value of the tail length of a randomly chosen individual) is greater, and she does an experiment to investigate this. She has the tail lengths y_i measured for a random sample of 10 individuals, and obtains the following results (in cm):

y_i	32.8	36.8	30.9	34.0	38.2	33.4	21.0	33.7	34.6	26.2

It is given that $\bar{y} = 32.16$ and $\sum(y_i - \bar{y})^2 = 233.324$.

a) Perform a test to investigate whether the expected tail length is greater than 30 cm. Assume that tail length is normally distributed. Use significance level $\alpha = 0.05$.

b) Find a 95% confidence interval for the expected tail length.

Also the latitude x_i where the animal stayed when the tail was measured war recorded. The table shows latitude and tail length for each animal:

x_i	63.9	61.5	64.8	65.5	59.0	58.5	68.5	66.0	66.0	66.8
y_i	32.8	36.8	30.9	34.0	38.2	33.4	21.0	33.7	34.6	26.2

Assume a linear regression model, where latitude is explanatory variable and tail length response variable. The biologist has a suspicion that the tail length decreases with incresing latitude.

It is given that $\bar{x} = 64.05$, $\sum(x_i - \bar{x})^2 = 100.465$, $\sum(x_i - \bar{x})(y_i - \bar{y}) = -105.88$.

c) Estimate the regression line (find $\hat{\beta}_0$ and $\hat{\beta}_1$). Perform a test to investigate the biologist's suspicion. Use significance level 0.05. Use that $\sum(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 121.737$.

Problem 3

Let $X_1, ..., X_n$ be a random sample from a normal distribution with expectation μ_x and variance σ^2, and $Y_1, ..., Y_m$ be a random sample from a normal distribution with expectation μ_y and variance $k\sigma^2$ Assume that X-s og Y-s all are independent. Parameters μ_x and μ_y are unknown and we have to estimate and to construct a confidence interval for the difference $\mu_x - \mu_y$.

We use $\bar{X}$ and $\bar{Y}$ as estimators of μ_x and μ_y respectively.

a) Argue that $\bar{X} - \bar{Y}$ has a normal distribution and show that

$$\mathrm{E}(\bar{X} - \bar{Y}) = \mu_x - \mu_y, \quad \mathrm{Var}(\bar{\mathrm{X}} - \bar{\mathrm{Y}}) = \sigma^2 \left(\frac{1}{n} + \frac{k}{m} \right).$$

b) Assume in this point that both parameters σ^2 and k are known. Construct an $(1 - \alpha)$-confidence interval for the difference $\mu_x - \mu_y$.

In the rest of the problem, we assume that the parameter σ^2 is unknown, while the parameter k is known.

Let S_x^2 and S_y^2 be empirical variances of X-ene and Y-ene, i.e.

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2, \quad S_y^2 = \frac{1}{m-1} \sum_{i=1}^{m} (Y_i - \bar{Y})^2.$$

It is known that

$$\frac{(n-1)S_x^2}{\sigma^2} \sim \chi^2_{n-1}, \quad \frac{(m-1)S_y^2}{k\sigma^2} \sim \chi^2_{m-1},$$

and that S_x^2 and $\bar{X}$ are independent, and S_y^2 and $\bar{Y}$ are independent.

c) Show that

$$S_p^2 = \frac{n-1}{n+m-2} S_x^2 + \frac{m-1}{k(n+m-2)} S_y^2$$

is an unbiased estimator of σ^2 and that

$$(n+m-2)S_p^2/\sigma^2 \sim \chi^2_{n+m-2}.$$

Show further that

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{S_p^2(\frac{1}{n} + \frac{k}{m})}}$$

has the Student t-distribution with $n + m - 2$ degrees of freedom.

Specify which properties and connections between various types of distributions you use and explain why necessary conditions are satisfied.

d) Construct a $(1 - \alpha)$-confidence interval for the difference $\mu_x - \mu_y$.

Assume that you can take part in the planning of the experiment. The total number of measurements, $N = n + m$, is given (is fixed due to economical reason) but you can choose n and m. How n and m have to be choosen to make the expected length of the confidence interval minimal?

Problem 4

Usually March is colder than April in Norway. Let X be the average temperature in March and Y the average temperature in April at Værnes a randomly chosen year, both measured in °C. Assume that X has the normal distribution with mean (expected value) μ_X and variance σ^2, and that Y has the normal distribution with mean μ_Y and variance σ^2. The average temperature in °C at Værnes for the years 2001-2012 was as follows:

	2001	2002	2003	2004	2005	2006
x_i (March)	-2.5	0.5	3.3	2.6	-0.7	-4.6
y_i (April)	4.1	7.2	5.0	7.9	5.8	4.9

	2007	2008	2009	2010	2011	2012
x_i (March)	3.3	0.8	1.9	-0.5	1.2	3.8
y_i (April)	5.0	5.9	6.9	4.8	6.7	3.2

It is given that

$$\sum_{i=1}^{12} x_i = 9.10,$$

$$\sum_{i=1}^{12} y_i = 67.40,$$

$$\sum_{i=1}^{12} x_i^2 = 77.07,$$

$$\sum_{i=1}^{12} y_i^2 = 399.30,$$

$$\sum_{i=1}^{12} (x_i - y_i)^2 = 364.53,$$

where $i = 1$ stands for year 2001, $i = 2$ for 2002 etc.

a) Assume that the March temperatures are independent from year to year, and the April temperatures are independent from year to year. We want to use hypothesis testing to try to show that the difference between expected average temperatures in April and March is less than 5°C. We can either use a two-sample test or a paired test. Which would you choose? Argue your choice, and perform the test you choose. Use significance level $\alpha = 0.05$. σ^2 is unknown.

Problem 5

We have two random variables X and Y. Let X have variance $\text{Var}(X) = 5$, and Y have variance $\text{Var}(Y) = 9$. Further, the covariance between X and Y is $\text{Cov}(X, Y) = 1$.

a) Calculate $\text{Cov}(2X + Y, X - Y)$.